

Abhinav Yedla

Washington, DC · 301-820-2727 · abhinavyedla02@gmail.com · abhinavyedla.com · linkedin.com/in/ayedla · github.com/abhinavyedla02

EXPERIENCE

Booz Allen Hamilton, Software Engineer II

July 2024 – Present

AI Solutions Architect: OTCnet AI Initiative, U.S. Treasury Bureau of the Fiscal Service

- **Created and now lead the AI initiative** on OTCnet, Treasury's ~\$150B/yr government-wide deposit-processing platform (96+ agencies): identified the use cases, defended them **directly to Fiscal Service**, and won the buy-in to build the **six-person AI team** that now owns the roadmap.
- Built an automated Section 508 accessibility triage system **running in CI/CD on every release**, using an Azure OpenAI Gov pipeline to classify severity, explain findings in plain language, and draft remediation. **Cuts review time ~40%** and catches violations manual review missed.
- Enabled any OTCnet workflow to use commercial frontier LLMs **without exposing controlled data**: built a tool that reversibly remaps identifiers out and back, now **adopted across requirements, security testing, DevOps, and PMO**.
- **Developed an NLQ service** over OTCnet's federal financial data so non-technical staff can query in plain language, with role-scoped access hardened by **AST-level validation and canary-row adversarial evals**.
- Built LLM drafting systems for security compliance artifacts: **SSP control narratives spanning 1,300+ controls** across NIST 800-53 and related frameworks, with per-claim citations and gap flagging where evidence falls short, and **Security Impact Analyses** drafted from de-identified release requirements across a draft pass and a verify pass.
- Scoped an **AI POC spanning four use cases** (NLQ, Insights & Trends, a customer-service chatbot, and an internal AI-in-SDLC track) and **won Fiscal Service approval** to move each into development.
- **Developing an MCP server** for OTCnet that exposes internal tools and live systems to AI agents, weighing governance and security tradeoffs in a regulated environment.

Department of Education, FAFSA AI Assistant & Cloud Infrastructure

- Built an **AI assistant** for FAFSA applicants that reads the user's in-progress form, fills it out through conversation, and flags aid opportunities they might otherwise miss, with multilingual and voice support.
- Built declarative **IaC** across containerized microservices for DoE workloads, contributing to a Kubernetes cloud migration.

Booz Allen Hamilton, Software Engineering Intern

May – Aug 2023

- Built an **automated federal procurement platform** to surface agency spending patterns via USASpending.
- Rebuilt Booz Allen's TFX internal platform to support scalable vendor onboarding.

PROJECTS

VTaaS: Video Transcoding as a Service · vtaas.abhinavyedla.com

2025 – 2026

- **Built and operate** a production async video-transcoding service on AWS **solo, end-to-end**: queue-decoupled background workers keep the API responsive through long-running transcodes.
- Designed an on-demand provisioning system **cutting AWS cost 99% (from \$58 to \$0.50/mo)**, resuming compute in ~4 minutes with 60-minute idle auto-hibernate.

2llm: AI Memory Migration Tool · 2llm.app

2026

- Converts AI memory exports between LLM providers in **under 5 minutes** instead of a manual rebuild.
- **BYOK with zero server-side key storage**, with memory kept session-only so uploaded conversations are never persisted.
- Added a managed conversion tier with Stripe payments; **75 early users** validating cross-provider migration demand.

SKILLS

AI/ML & GenAI: RAG, hybrid keyword + vector retrieval, vector databases; Anthropic API, OpenAI API, Azure OpenAI; agentic patterns (MCP, multi-model routing, human-escalation paths); LangChain, prompt engineering, evaluation-framework design

Languages: Python, TypeScript, JavaScript, Java, C/C++, SQL, Bash

Cloud & Infrastructure: AWS (ECS Fargate, S3, SQS, ECR, IAM, CloudWatch, ALB, SSM), Azure (OpenAI Service, AI Search, Functions, DevOps, Load Testing), Docker, Kubernetes, Terraform, AWS CDK, Vercel, Neon, Upstash

Frameworks & Libraries: React, Next.js, NestJS, Node.js, Flask, FastAPI, Prisma, REST APIs

Databases: PostgreSQL, MongoDB, Redis, DynamoDB

EDUCATION

Georgia Institute of Technology, M.S. Computer Science (OMSCS)

Starting Fall 2026

University of Maryland, College Park, B.S. Computer Science

Spring 2024

AWS Certified Solutions Architect – Associate (2025) · Oracle Certified Associate, Java SE 8 (2025)